Southwest Fisheries Center Administrative Report No. 13H, 197

BOUNDS BY DOMINANCE FOR FIXED-WEIGHT AGGREGATED MARKOV DECISION PROCESSES

Roy Mendelssohn
Southwest Fisheries Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
Honolulu, Hawaii 96812

ABSTRACT

Bounds by dominance for aggregated Markov decision processes are developed. These bounds only require the evaluation of K columns as compared to $N \cdot M$ columns, $K < N \cdot M$, and $N \cdot M$ is the total number of actions in all states combined. Separate bounds are given when only actions within a state are aggregated, when states are aggregated only in the rows of the linear programming solution, and for the general aggregate problem. A procedure to tighten the bounds is discussed, and a numerical example is given.

2. NOTATION

In each period, a state x from a finite set of states is observed, a feasible action y is taken, and a transistion is made to state j with probability $p(x, j:y)$. When state x is observed and decision y is taken, an expected return of $g(x, y)$ is received. The return is discounted through time by a discount factor α, $0 \leq \alpha < 1$.

It is assumed that there are N states, N finite, and without loss of generality, the same M feasible actions per state. It is desired to maximize the expected total return over an infinite planning horizon, and it is assumed that the one-period returns are uniformly bounded from below. This is sufficient to assume that the returns are nonnegative. Given these assumptions, the well known LP solution to this MDP is (d'Epenoux [1])

$$z^* = \text{maximize} \sum_{x=1}^{N} \sum_{y=1}^{M} g(x, y)\, u(x, y)$$

$$\text{subject to} \quad \sum_{x=1}^{N} \sum_{y=1}^{M} \Big(\delta_{xj} - \alpha p(x, j:y)\Big) u(x, y) \leq 1 \qquad j = 1, \ldots, N \tag{2.1}$$

$$u(x, y) \geq 0 \qquad x = 1, \ldots, N;\ y = 1, \ldots, M$$

where $\delta_{xj} = \begin{cases} 1 & \text{if } x = j \\ 0 & \text{if } x \neq j \end{cases}$ and the inequalities in (2.1) are justified by the nonnegativity of the expected return. Optimal solutions to (2.1) are denoted by u^*, and optimal dual variables by v^*.

A major deterrent to real world use of Markov decision processes (MDPs) is the large problem size that may arise. For example, some fisheries models may have a seven-dimensional state vector. If only 10 grid points are used in each dimension, the resulting MDP would have 10 million states, and each policy would have a transistion matrix too large to calculate even one iteration of successive approximations. Recently, several authors (Hinderer and Hübner [2], White [6], Whitt [7, 8]) have studied the following approach for approximating MDPs: aggregate the states and actions of the original model into a smaller model, solve this aggregated or reduced model, extend the solution of the aggregated model to the original model, and calculate bounds on the loss in value from using this suboptimal policy. These bounds are a posteriori bounds, and require at least the computational equivalent of one iteration of successive approximations on the original model. For problems of similar size as the example above, it would not be practical to calculate the bounds.

In this paper, some results of Zipkin [9, 10] for aggregated linear programming problems (LPs) are used to find a family of bounds that require less computational effort but are looser bounds. These bounds allow the modeler to solve a reduced problem of one size and calculate the bounds on an aggregate problem of a different size, thus allowing the modeler some choice between tightness of the bound and computational effort. A method to improve these "bounds by dominance" is given, and a numerical example is presented where the improved bounds by dominance are almost as tight as the best improved bound, but the computational effort was significantly less.

The notation and the method of aggregation and disaggregation follows Zipkin [9, 10]. The reader is referred to [9, 10] for more details. Let p be the $N \times N \cdot M$ constraint matrix in (2.1), and let G be the $N \cdot M$-vector of objective function coefficients in (2.1). Let $\bar{\sigma} = \{\bar{S}_k;\ k = 1, \ldots, \bar{K}\}$ be a partition of $\{1, \ldots, N \cdot M\}$, and let $\bar{\rho} = \{\bar{R}_\ell:\ \ell = 1, \ldots, \bar{L}\}$ be a partition of $\{1, \ldots, N\}$. Let p^k be the submatrix of p consisting of those columns whose indices are in $\bar{S}_k$, p_ℓ the submatrix whose rows have indices in $\bar{R}_\ell$, and p_ℓ^k the submatrix whose columns and rows have indices in $\bar{S}_k$ and $\bar{R}_\ell$. Let G^k be the components of G whose indices are in $\bar{S}_k$. Let c^k and f^ℓ be nonnegative vectors of appropriate lengths, whose components sum to unity. Then a fixed-weight column aggregated problem, fixed-weight row aggregated problem and doubly aggregated problem are defined by the following transformations:

COLUMN	ROW	DOUBLY	
$\bar{G}^k = G^k c^k$	$\underline{p}_\ell = f^\ell p_\ell$	$\tilde{G}^k = G^k c^k$	(2.2)
$\bar{p}^k = p^k c^k$		$\tilde{p}_{k\ell} = f^\ell p_\ell^k c^k$	
		$\tilde{p} = (\tilde{p}_{k\ell})$	
		$\tilde{b} = f^\ell \underset{\sim}{1}$	

These transformations are equivalent to the projection operators in Whitt [7].

The new variables for the column and doubly aggregated problems are denoted by μ, and the new dual variables by ν. Optimal values for the aggregated problem are denoted $\bar{\mu}$, $\bar{\nu}$. For example, the LP for the doubly aggregated problem is:

$$\tilde{z} = \text{maximize} \sum_{k=1}^{K} \tilde{G}^k \mu^k \qquad (2.3)$$

$$\text{s.t.} \quad \tilde{p}\mu \leq \tilde{b}$$

$$\mu \geq 0$$

The solution of the aggregated problem is disaggregated using the identical transformation; for example, if rows are aggregated, then $\bar{\nu}$ is disaggregated as $\tilde{v}^\ell = \nu_\ell f^\ell$, where $\tilde{v}^\ell$ is a vector with as many elements as in $\bar{R}_\ell$. The disaggregated variables will be denoted $\tilde{u}$, $\tilde{v}$. These transformations are Whitt's [7] extension operators.

Let $\sigma = \{S_k: k = 1, \ldots, K\}$ and $\rho = \{R_\ell: \ell = 1, \ldots, L\}$ be any two partitions of $\{1, \ldots, N \cdot M\}$ and $\{1, \ldots, N\}$ perhaps differing from $\bar{\sigma}$, $\bar{\rho}$. Let $\{d_j\}$ and $\{e_i\}$ be known positive numbers, and $\{w_k\}$ and $\{q_\ell\}$ known nonnegative numbers.

<u>Theorem</u> (Zipkin [10]): <u>If it is known a priori that:</u>

$$\sum_{j \in S_k} d_j u_j^* \leq w_k \qquad k = 1, \ldots, K \qquad (2.4)$$

$$\sum_{i \in R_\ell} e_i v_i^* \leq q_\ell \qquad \ell = 1, \ldots, L$$

<u>then</u> $\tilde{z} - \varepsilon_a^- \leq z^* \leq \tilde{z} + \varepsilon_a^+$

<u>where</u>

$$\varepsilon_a^+ = \sum_{k=1}^{K} \left\{ \max_{j \in S_k} \left[(G^j - \tilde{v}p^j)/d_j \right] \right\}^+ w_k \qquad (2.5a)$$

$$\varepsilon_a^- = \sum_{\ell=1}^{L} \left\{ \max_{i \in R_\ell} \left[(p_i \tilde{u} - 1)/e_i \right] \right\}^+ q_\ell \qquad (2.5b)$$

□

Three cases are treated separately in the next section. In the first case, only actions within a state can be aggregated. These partitions of $\{1, \ldots, M\}$ will be denoted by $\bar{\sigma}_x = \{\bar{S}_{xk}: k = 1, \ldots, \bar{K}\}$ with $\bar{\sigma} = \underset{x}{\cup}\bar{\sigma}_x$. The second case has columns aggregated as in the first case, but states may also be aggregated in the rows. The third case is the more general doubly aggregated problem.

For any group of partitions $\sigma_x = \{S_{xk}: k = 1, \ldots, K\}$ define $G^k(x) \geq \max_{y \in S_{xk}} G(x, y)$ and $p^k(x, j) \geq \max_{y \in S_{xk}} p(x, j:y)$. Let $\{d_{xk}\}$, $\{w_{xk}\}$ be the obvious redefinitions of $\{d_k\}$ and $\{w_k\}$. The results of this paper utilize the fact that $G^k(x)$ and $p^k(x, j)$ can be used to simplify (2.5). The original bounds (2.5), like the bounds in Whitt [7], at least require evaluating one iteration of successive approximation. The new bounds do not, but with an obvious loss of value.

3. RESULTS

Only the proof for the first theorem will be given in detail, since the proofs of the other theorems are similar. The main result used is the fact that "max $\Sigma \leq \Sigma$ max." Theorem 3.1 gives the equivalent bound by dominance to (2.5a).

Theorem 3.1 *If only actions within a state are aggregated, let z^* be the value of a solution to the original LP and $\tilde{z}$ the value of a solution to the aggregated LP, then:*

$$\tilde{z} \leq z^* \leq \tilde{z} + \varepsilon_+$$

<u>where</u> $$\varepsilon_+ = \sum_{x=1}^{N} \sum_{k=1}^{K} \left\{ \left[\left(G^k(x) + \alpha \Sigma p^k(x, j)\tilde{v}_j - \tilde{v}_x \right) \Big/ d_{xk} \right] \right\}^+ w_{xk} \qquad (3.1)$$

<u>Proof</u>. The theorem would be true if for each of the N·K columns in the aggregate problem:

$$G^k(x) + \alpha \Sigma p^k(x, j)\tilde{v}_j - \tilde{v}_x \geq \max_{j \in S_{xk}} \left\{ G(x, y) + \alpha \Sigma p(x, j{:}y)\tilde{v}_j - \tilde{v}_x \right\}$$

However, the result is immediate from the definition of $G^k(x)$, $p^k(x, j)$.

□

The bound (3.1) requires the evaluation of N • K inner products between two N-vectors, while (2.5a) requires the evaluation of N • M inner products between two N-vectors. For N large and K<<M, the computational savings is substantial. A likely procedure would be to choose $\bar{K} \leq$ K<<M. Note also that the values of $G^k(x)$, $p^k(x, j)$ need not be found by sorting through the constraint matrix, but can be well chosen values based on knowledge of the problem.

If states are aggregated together only in the rows, then the upper bound in theorem 3.1 remains valid. However, the lower bound must be altered. Define $p_\ell(x{:}y) = \min_{j \in R_\ell} p(x, j{:}y)$, $e_\ell = \min_{j \in R_\ell} e_j$ and for each $j \in R_\ell$, $\tilde{u}(j) = \sum_{y=1}^{M} \tilde{u}(j, y)$. For most usefully aggregated problems, only one value of $\tilde{u}(j, y)$ will be nonzero. Then let $\tilde{u}_\ell = \max_{j \in R_\ell} \tilde{u}(j)$.

<u>Theorem 3.2</u> <u>The assumptions of theorem 3.1, plus states being aggregated in the rows only, imply</u>

$$\tilde{z} - \varepsilon_- \leq z^* \leq \tilde{z} + \varepsilon_+$$

<u>where</u> $$\varepsilon_- = \sum_{\ell=1}^{L} \left\{ \left[\sum_{x=1}^{N} \sum_{g=1}^{M} \left[u_\ell - \alpha p_\ell(x{:}y)\tilde{u}(x, y) - 1 \right] \Big/ e_\ell \right] \right\}^+ q_\ell \qquad (3.2)$$

□

The bound in theorem 3.2 involves evaluating L inner products between two N•M-vectors, while (2.5b) requires evaluating N inner products between two N•M-vectors. For N•M large and L<<N, this is a substantial saving. Both theorems 3.1 and 3.2 require that the optimal solution or optimal dual variables be disaggregated, and that the upper (resp. lower) bounds $G^k(x)$, $p^k(x, j)$ $\Big($resp. $p_\ell(x{:}y)\Big)$ be based on the original problem. In fact, it is not necessary to totally disaggregate the final solution of the aggregate LP, and the bounds can be based solely on the partially aggregated tableau. This is proven for the upper bound in theorem 3.2. The lower bound is proven similarly. Define $\bar{p}(x, \ell{:}y)$ as the aggregate probability of going from state x to the new row-aggregate state ℓ when action y has been taken. And let $\bar{p}^k(x,\ell) \geq \max_{y \in S_{xk}} \bar{p}(x, \ell{:}y)$.

<u>Theorem 3.3</u> <u>The assumptions of theorem 3.2 imply</u> $z^* \leq \tilde{z} + \varepsilon_+^2$

<u>where</u> $$\varepsilon_+^2 = \sum_{x=1}^{N} \sum_{k=1}^{K} \left\{ \left[G^k(x) + \alpha \sum_\ell \bar{p}^k(x, \ell)\bar{\nu}_\ell - \tilde{\nu}_x \right] \Big/ d_{xk} \right\}^+ w_{xk} \qquad (3.3)$$

<u>Proof</u>. From (2.5a), $\sum_{k=1}^{K}\left[g(x, y) + \alpha \sum_j p(x, j:y)\tilde{v}_j - \tilde{v}_x\right]$

$$= \sum_{k=1}^{K}\left[g(x, y) + \alpha \sum_{\ell} \sum_{j \in R_\ell} p(x, j:y)\ (\bar{\nu}_\ell f_\ell^j) - \tilde{v}_x\right]$$

$$= \sum_{k=1}^{K}\left[g(x, y) + \alpha \sum_{\ell} \bar{\nu}_\ell \sum_{j \in R_\ell} p(x, j:y) f_\ell^j - \tilde{v}_x\right]$$

$$= \sum_{k=1}^{K}\left[g(x, y) + \alpha \sum_{\ell} \bar{p}(x, \ell:y)\bar{\nu}_\ell - \tilde{v}_x\right]$$

The rest of the proof is the same as the proof of theorem 3.1.

□

The bound in theorem (3.3) requires evaluating $N \cdot K$ inner products between two L-vectors instead of two N-vectors. For L<<N, the increased savings in computational effort is significant. If states are to be aggregated in the columns also, it is reasonable to assume that an aggregate would be formed by aggregating the same actions between states to produce a new aggregate state with M action but with aggregate transition probabilities, and then to further aggregate the actions within the aggregate state. Let $\sigma = \{S_k : k = 1, \ldots, K\}$ be a partition of $\{1, \ldots, N\}$, and S_k index the states that have been (column) aggregated together. Let $\tilde{v}_k = \underset{x \in S_k}{\text{minimum}}\ \tilde{v}_x$. Then theorem 3.3 extends to the more general aggregation scheme by substituting $\tilde{v}_k$ for $\tilde{v}_x$ in (3.3)

4. EXTENSIONS AND EXAMPLES

In Mendelssohn [4], a method of Kallio [3] is used to improve all of the bounds discussed in this paper. This is done by including the constraints (2.4) directly into the LP (2.1), and also adding the additional constraint that the vector of variables u must lie in the set $U = \{u \mid u = \theta u, \theta \in R\}$. This additional constraint produces a parametric family of values $z^*(\theta)$ and a parametric family of bounds. At $\theta = 1$, the bounds are equivalent to the original bounds. A search procedure is used to find the value of θ that produces the smallest bound.

It has been shown mathematically that the bounds by dominance require significantly less computation; it would be hoped that this is without a concomitant loss in the tightness of the bound. As an example, a model suggested by Mathews [11] for salmon runs in the Naknek River is analyzed. The "original" problem has a state space of 31 points lying between 0 and 7 million fish. The aggregate
Table I, II problem and its solution are presented in Table I, and Table II gives the variable definitions, the stochastic transistion function, and four different error bounds calculated from this example.

As can be seen in Table II, the bounds by dominance are not tight at all, but the improved bounds by dominance are almost as tight as the best available bound. One example does not imply that improved bounds by dominance will always, or even often, provide tight bounds. However, the example shows that they <u>can</u> provide tight bounds while requiring significantly less computational effort; further computational work needs to be performed on real life models to study the behavior of these bounds.

A more intriguing question, which has not been examined at all, is if any of the bounds are monotone in the sense that a tighter bound implies that the expected value of the extended policy on the original model is greater. Such monotonicity properties would provide clear guidelines for modeling. Finally, the Russian literature over the last decade or so has studied processes that iteratively aggregate using the extended solution from the previous aggregate problem to reaggregate the model (see Vakhutinsky, Dudkin, and Ryvkin [5] and references therein). A combination of iterative aggregation combined with bounds to guide the reaggregation process would appear to be a fruitful area of future research.

ACKNOWLEDGMENTS

I am grateful to Profs. M. J. Sobel and P. Zipkin who shared with me many of their unpublished ideas for approximating MDP's; this paper is an outgrowth of those discussions. The associate editor and a referee provided numerous suggestions that greatly improved the content of this paper.

REFERENCES

1. D'EPENOUX, G., "A Probabilistic Production and Inventory Problem," Management Sci. 10, 98-108 (1963).

2. HINDERER, K. and G. HÜBNER, "Recent Results on Finite-Stage Stochastic Dynamic Programs," Paper Presented at the 41st Session, International Institute of Statistics, New Delhi, December 1977.

3. KALLIO, M., "Computing Bounds for the Optimal Value in Linear Programming," NRLQ. 24, 301-308 (1977).

4. MENDELSSOHN, R., "Improved Bounds for Aggregated Linear Programs," Submitted for Publication (1979).

5. VAKHUTINSKY, I. Y., L. M. DUDKIN, and A. A. RYVKIN, "Iterative Aggregation - A New Approach to the Solution of Large Scale Problems," Econometrica 47, 821-841 (1979).

6. WHITE, D. J., "Finite State Approximations for Denumerable State Infinite Horizon Discounted Markov Decision Processes," Note #43, Department of Decision Theory, University of Manchester, 1977.

7. WHITT, W. "Approximations of Dynamic Programs, I," Math. Opns. Res. 3, 231-243 (1978).

8. WHITT, W. "Approximations of Dynamic Programs, II," Math. Opns. Res. 4, 179-185 (1979).

9. ZIPKIN, P. "Bounds on the Effect of Aggregating Variables in Linear Programs," Research Paper No. 211, Graduate School of Buiness, Columbia University, 1977. (To appear, Operations Research.)

10. ZIPKIN, P. "Bounds for Row-Aggregation in Linear Programming," Research Paper No. 105A, Graduate School of Business, Columbia University, 1978.

11. MATHEWS, S. B. "The Economic Consequences of Forecasting Sockeye Salmon Runs to Bristol Bay, Alaska: A Computer Simulation Study of the Potential Benefits to a Salmon Canning Industry from Accurate Forecasts of the Runs," Ph.D. Dissertation, University of Washington, Seattle, 1967.

TABLE I

Aggregated LP for Naknek salmon

Partition	Variables in the partition	Weighting for aggregation	Upper bound P_k
1	$\{u_x^0, u_x^{0.125}; x \in X\}$	Zero on u_x^0 1/30 on $u_x^{0.125}$	35
2	$\{u_x^{0.25}; x \in X\}$	1/29 on each column	10
3	$\{u_x^{0.375}; x \in X\}$	1/28 on each column	10
4	$\{u_x^{0.5}; x \in X\}$	1/27 on each column	10
5	$\{u_x^{0.625}; x \in X\}$	1/26 on each column	10
6	$\{u_x^{0.75}; x \in X\}$	1/25 on each column	1,000
7	$\{u_x^{0.875}; x \in X\}$	1/24 on each column	10
8	$\{u_x^{1}; x \in X\}$	1/23 on each column	10
9	$\{u_x^{1.125}; x \in X\}$ and	1/22 on each column	10
	$\{u_x^y; 9 \geq x \geq 1.125;$ $1.25 \leq y \leq x; x, y \in X\}$	Zero on each column	

$$G^k(x, y) = \{9, 8.75, 8.625, 8.5, 8.375, 8.25, 8.125, 8, 7.875\}$$

$$p^k = \{1, 0.7, 0.7, 0.41, 0.7, 0.7, 0.7, 0.7, 0.7, 0.2, 0.7, 0.7, 0.7, 0.7, 0.7, 0.25, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7\} \text{ for all } k = 1, \ldots, K.$$

The optimal dual variables for the aggregated LP are:

$$\bar{f} = \{0, 0, 321.0216, 92.4958, 0, 0, 0, 0, 0, 179.4263, 0, 0, 0, 0, 0, 31.3039, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0\}$$

TABLE II

Example of the Different Bounds

Model: Salmon runs in the Naknek River:

x_t = Number of fish that return to river at time t

y_t = Number of fish released to spawn at end of time t

$x_t - y_t$ = Catch = One-period return

$x_{t+1} = 6.727 \exp(d)\, y_t \exp\{-0.859\, y_t\}$

$d \sim N(0, 0.1444)$

Original problem has 31 states, $z^* = 1918.1672$

Regular bounds	Bounds by dominance	Improved bounds	Improved bounds by dominance
$624.25 \le z^* \le 14{,}849.09$	$624.25 \le z^* \le 347{,}663$	$624.25 \le z^* \le 9{,}051.62$	$624.25 \le z^* \le 9{,}147.00$